Yale iGEM 2015 Research Summary

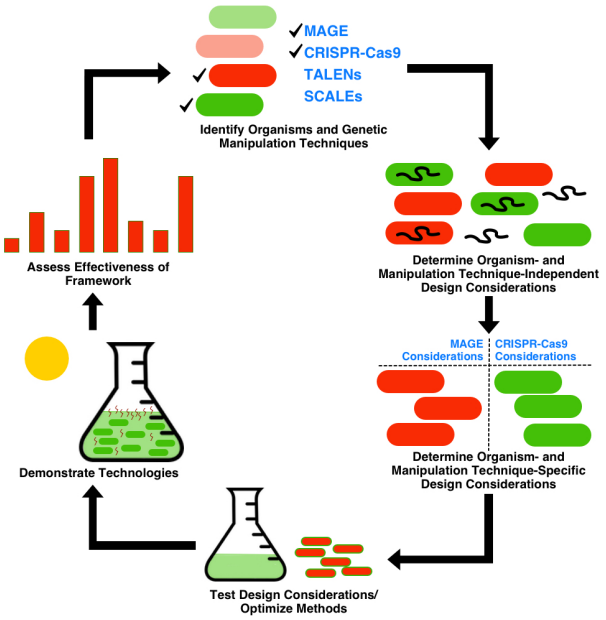
Developing a Framework for the Genetic Manipulation of Non-Model and Environmentally Significant Microbes

Colin Hemez[1], Lionel Jin[1], Danny Keller[1], Dan Shapiro[1], Jessica Tantivit[1], Erin Wang[1], Holly Zhou[1], Joe Lanzone[2], Andrew Saydjari[2], Ariel Hernandez-Leyva[3], Natalie Ma[4], Jaymin Patel[4], Corey Erik Perez[5], Stephen Dellaporta[6], Farren J. Isaacs[6]

[1] Undergraduate wet-lab researcher, Yale University iGEM. These authors contributed equally to research.
[2] Undergraduate modeling team, Yale University iGEM
[3] Undergraduate Research coordinator, Yale University iGEM
[4] Graduate student mentor, Yale University Isaacs Lab
[5] Graduate student mentor, Yale University Crawford Lab
[6] Faculty Advisor, Yale University iGEM

Project Abstract

We focused on establishing a framework to implement genetic manipulation techniques—specifically, multiplex automated genome engineering (MAGE) and CRISPR-Cas9 systems—into non-model, environmentally significant microbes using standard biological parts. The framework involves two components: (1) propagation and selection of cultures and (2) manipulation of cell genomes by MAGE and/or CRISPR. We identified design considerations for both components of the framework, and experimentally validated propagation and selection considerations using cyanobacterial strain *Synechococcus* sp. PCC 7002 (a fast-growing cyanobacterium capable of lipid biofuel production) and *Sinorhizobium tropici* CIAT (a nitrogen-fixing rhizobium which forms root nodules in legume plants). We then developed a workflow for the design, construction, and testing of MAGE and CRISPR technologies into non-model prokaryotes. The insights we gained from validating the propagation component of our workflow will improve the versatility and robustness of our framework and will inform the development of tools for genetic manipulation in other non-model organisms.



Introduction

Our iGEM research project involves developing a framework for the implementation of genetic manipulation techniques—specifically, multiplex automated genome engineering (MAGE) and CRISPR-Cas9 systems—into non-model, environmentally significant microbes. MAGE was developed as a rapid, high efficiency tool for increasing the genetic diversity of a cell population at targeted loci within the genome, and has so far been ported into the model organism *Escherichia coli* and a few other members of the family Enterobacteriaceae (unpublished data). CRISPR (clustered regularly interspaced short palindromic repeats)/Cas systems, based on prokaryotic adaptive immunity mechanisms, have emerged as a powerful cleavage-based genome editing technique (Cong et al. 2013, Hsu et al. 2014).

The organisms we used to develop our framework are *Rhizobium tropici* CIAT 899 and *Synechococcus* sp. PCC 7002. *Rhizobium tropici* CIAT 899 (hereafter CIAT899) is a nitrogen-fixing bacterium capable of forming root nodules with certain legume-producing *Leucaena* and *Phaseolus* trees (Martinez-Romero et al). *Synechococcus* sp. PCC 7002 (hereafter PCC7002) is a fast-growing marine cyanobacterium capable of photosynthesis and free fatty acid (FFA) production (Ruffing 2014).

We envision numerous potential applications for MAGE, CRISPR-Cas9, and other genetic manipulation techniques in these organisms; for example, the nitrogen fixation mechanisms of CIAT899 could be modified to enable plant growth in

otherwise hostile environments, and the FFA biosynthesis pathway of PCC7002 could be optimized for the production of molecules that serve as precursors to lipid biofuels.

Methods

We divided the workflow for the implementation of MAGE and CRISPR-Cas9 into PCC7002 and CIAT899 into two sections: (1) propagation and selection of cultures and (2) manipulation of cell genomes using protocols specific for MAGE or CRISPR (Fig. 1). The first component of the workflow involves determining optimal growth, transformation, and selection protocols for each organism.

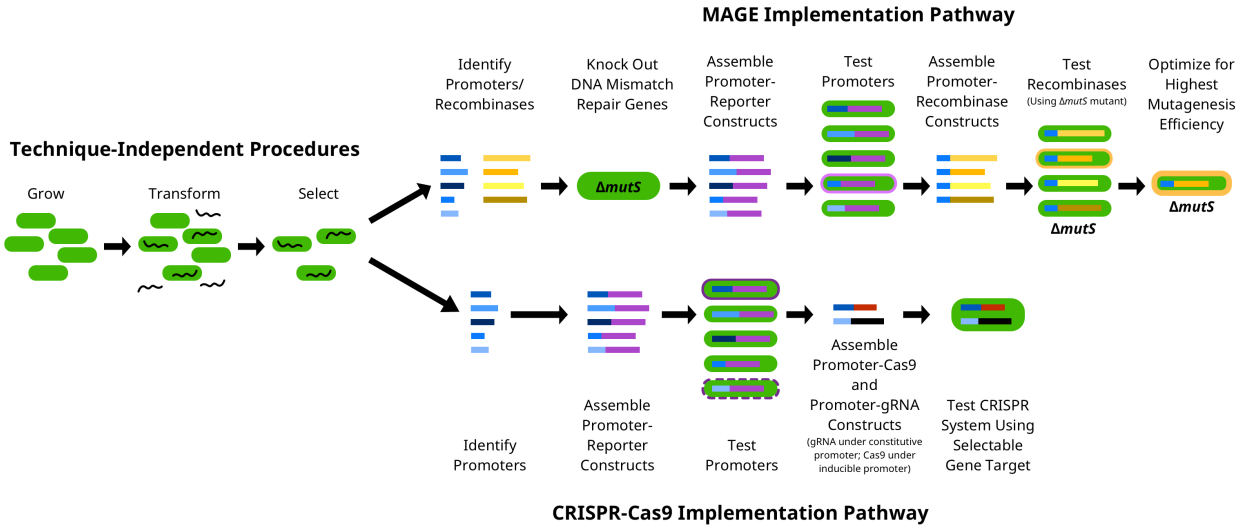


Figure 1: Methods used to implement MAGE and CRISPR-Cas9 systems into PCC7002 and CIAT899. The methods were divided into technique-independent and technique-specific protocols.

Technique-Independent Procedures

Liquid cultures of PCC7002 were grown at 38 ˚C and were supplemented with light (10,000 lux, 6400K) and CO_2. The media used was A+ supplemented with 20 µg/l vitamin B12. 10 mM glycerol was occasionally added to the media as it has been shown to lead to modest improvements in growth rate without shunting cells away from photosynthetic pathways (Ludwig and Bryant 2012). Liquid and solid cultures of CIAT899 were grown in either 50% Tryptic Soy Broth (TSB) (7.5g BD™ Tryptic Soy Broth in 500mL dH_2O) or Lysogeny Broth (LB) at 30 °C.

Multiple transformation methods were investigated for both PCC7002 and CIAT899, including natural competency, electroporation, conjugation, and chemical competency. Protocols were determined to be optimal for either organism based on simplicity and time needed to execute the procedure. PCC7002 cells were transformed via natural competency using a modified protocol from Frigaard et al. 2004. When transforming linear constructs for recombination with the PCC7002 genome, single colonies were screened for homozygosity by colony PCR. CIAT899 was made electrocompetent and transformed using electroporation (2.5V, 200 ohms, and 25 mF). The electroporated cultures were recovered for 4 hours and plated on antibiotic plates to select for cells that contain the vector of interest. We experimented with several broad host range vectors which could be used to transform gene constructs into PCC7002, CIAT899, and other non-model organisms in which this framework is employed to implement genetic manipulation techniques.

When performing transformations in PCC7002, the species' natural recombinogenecy was taken advantage of by transforming linear fragments with 1000 bp homology arms flanking either end of the insert of interest. The linear construct undergoes a double recombination event in some cells and incorporates into the genome.

Minimum inhibitory concentrations (MICs) for both PCC7002 and CIAT899 were determined experimentally using 96-well plate assays. A total of six antibiotics (carbenicillin, kanamycin, rifampicin, spectinomycin, streptomycin, and tetracycline) were tested at various concentrations for each organism. We considered multiple factors when deciding which antibiotic to use in subsequent genetic manipulation experiments, such as the value of the MIC, the availability of associated genetic resistance cassettes, and possible interactions between the antibiotic and the growth media (pH of the media, presence of phosphates in the media).

Technique-Specific Procedures

We identified promising inducible and constitutive promoters from the Registry of Standard Biological Parts and through genome searches of PCC7002 and CIAT899. Promoters native to our organisms were isolated by colony PCR. In order to identify candidate recombinases, we ran NCBI BLAST searches of the beta protein in the bacteriophage lambda recombineering cassette against cyanobacterium- and rhizobium-infecting phages. The identified rescombinases (beta-homologs) were synthesized as IDT™ gBlocks®.

Both the MAGE implementation pathway and the CRISPR-Cas9 implementation pathway require the construction of many DNA constructs from standard parts; thus, we sought a high-throughput method for assembling constructs from promoter, recombinase, and reporter gene fragments. We explored Gibson Assembly (GA, from New England Biolabs™), ligation-independent cloning (LIC), and exonuclease and ligation-independent cloning (ELIC). One of the main motivators for pursuing methods other than GA was its high cost of $12-16 USD per reaction.

In order to knock out the *mutS* gene in PCC7002 and CIAT899, we adapted FLP-FRT recombination protocols as previously described in E. coli and *Synechocystis* sp. PCC 6803, a strain of freshwater cyanobacteria (Datsenko and Wanner 2000, Tan et al. 2013). The mutS protein, involved in the DNA mismatch repair pathway, recognizes single base mutations (Modrich and Lahue 1996). Since such mutations are the basis of MAGE, allowing them to go unnoticed by the cell's proofreading machinery will increase the efficiency of each MAGE cycle. The FLP system, derived from the yeast 2µM plasmid, relies on a flpase recombinase to excise sequences flanked by two flpase recognition target (FRT) sites (Schweizer 2003). The process leaves behind a single FRT site scar. Our approach involved incorporating a kanamycin resistance cassette flanked by FRT sites into the mutS locus of each organism by natural homologous recombination. A flpase-containing plasmid (derived from pCP20) is then transformed into homozygous mutants. The flpase, induced by heat, excises the kanamycin cassette and leaves behind an FRT scar in place of the *mutS* gene. The plasmid is heat-curable and degrades shortly after the flpase is induced.

We used *mCit*, a yellow fluorescent protein (citrine) gene to characterize our promoters. Citrine was chosen over green fluorescent protein (GFP) because it does not emit near the autofluorescence range of PCC7002; testing promoters in PCC7002 with GFP could have lead to confounded results.

We also designed a method for testing the mutagenecy of an inducible recombinase system (MAGE) at multiple resolutions while simultaneously knocking out *mutS* (Fig. 2) A linear construct consisting of an FRT-flanked kanamycin cassette, an ampicillin resistance gene (*ampR*) with a premature stop codon, and an *mCit* gene with a premature stop codon (all flanked by 1000 bp homology arms) can be incorporated at the genomic *mutS* locus. Oligonucleotides targeting restored function of the premature stop codon of either the *ampR* cassette or the *mCit* gene can be transformed into mutants, along with recombinases under inducible promoters. Targeting either broken gene will yield a different dynamic range for assaying the mutation frequency per MAGE cycle—the ampR gene can be targeted to determine mutation frequencies on the order of 10^{-7} to 10^{-5}, while the mCit gene can be targeted to determine mutation frequencies from 10^{-4} to 10^{-1}.

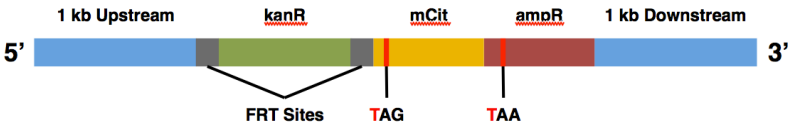
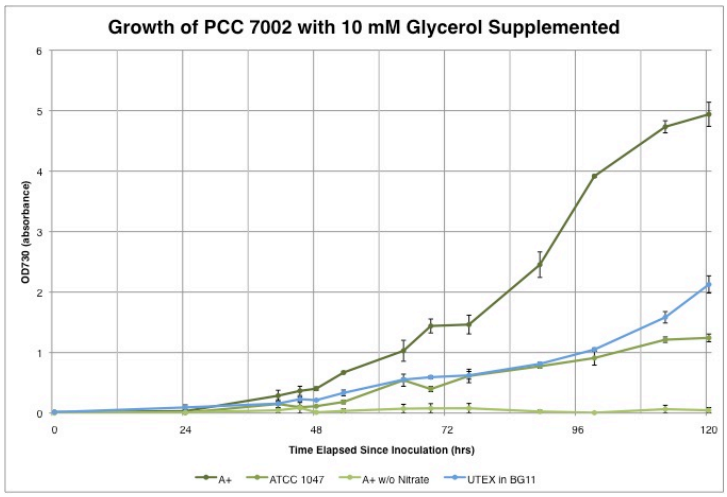
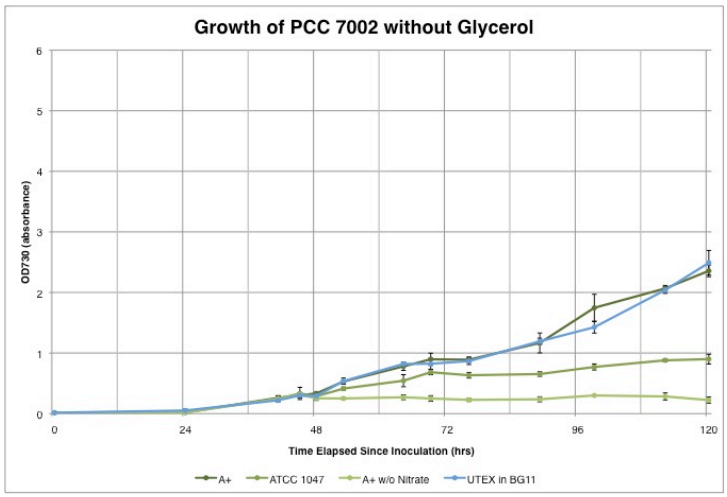


Figure 2: Schematic of the MAGE testing construct designed for PCC7002. The construct will be incorporated into the *mutS* locus of the PCC7002 genome by homologous recombination. The *mCit* and *ampR* genes contain premature stop codons caused by single-base mutations which will be targeted using oligonucleotides in MAGE cycling to restore gene function (i.e. the oligos will incorporate a T –> G mutation in *mCit* and a T –> C mutation in *ampR*).

Results

We managed to experimentally validate the technique-independent procedures (growth, selection, and transformation) of our framework for both PCC7002 and CIAT899. The growth rate of PCC7002 was found to be highly sensitive to growth conditions, particularly light saturation and CO_2 content of air. PCC7002 samples grown in CO_2-supplemented conditions grew significantly faster than samples grown in atmospheric conditions, and samples supplemented with 10mM glycerol grew to a much higher saturation than non-supplemented samples in the same period of time (Fig. 3a-b). The glycerol was also found to not shunt cells away from photosynthetic pathways, as populations grown in glycerol had the same chlorophyll content per cell as nonsupplemented populations (Fig. 3c). PCC7002 was found to be susceptible to all antibiotics tested around standard E. coli concentrations, though light-sensitive antibiotics (tetracycline and rifampicin) were less effective due to degradation in light-saturated growth chambers.





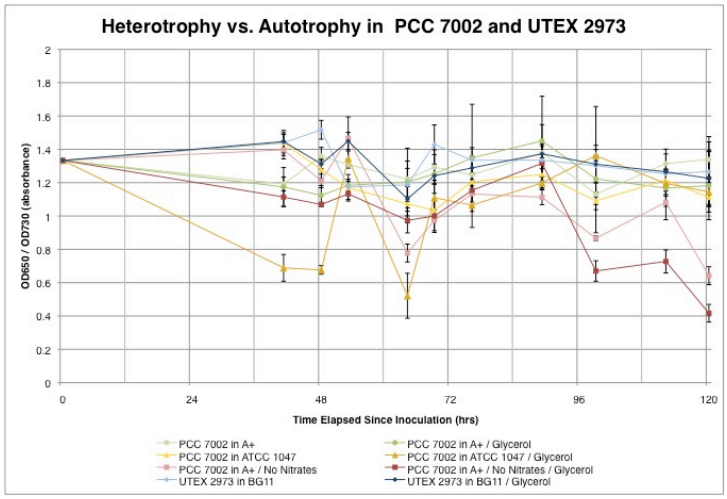


Figure 3: (a) Growth of PCC7002 (dark green) and UTEX 2973 (blue, another fast-growing strain of freshwater cyanobacterium) in 10mM glycerol-supplemented conditions, measured by OD 730nm. (b) Growth of PCC7002 and UTEX 2973 in media not supplemented with glycerol, shown on the same scale as (a). (c) Ratio of chlorophyll content (OD 650 nm) to cell saturation (OD 730 nm) for cell populations grown in glycerol-supplemented versus nonsupplemented conditions. PCC7002 grown in A+ medium is shown in green.

Using search tools that compare nucleotide sequences to organisms and viruses in NCBI and UniProtKB databases, nine λ beta and SPP1 35 homologues were identified that exist in Rhizobium species (Table 1). Our growth assay indicated that CIAT899 has a doubling time of 2.5 hours in 50% TSB media (Fig. 4). Of the antibiotics tested (Streptomycin, Carbenicillin, Kanamycin, Rifampicin, Spectomycin) we found CIAT899 has resistance to only Spectomycin, thus CIAT899 containing our broad host range plasmid can be selected for. Electroporation of our broad host range plasmid was successful in CIAT899.

When testing the expression of promoter-citrine constructs in Rhizobia, we found that the tac and Anderson Medium promoters resulted in the greatest fluorescence (Fig. 5).

We also successfully obtained *mutS* knockout genotypes of PCC7002 by transforming linear fragments for homologous recombination. Presumably due to the polyploidy of PCC7002 cells, however, we did not obtain homozygous Δ*mutS* mutants (Fig. 6); doing so will require several more rounds of growth and plating for single colonies.

Figure 4: Growth curve of CIAT899 and three strains of *Sinorhizobium meliloti* 1021, as measured by a 96-well plate reader. The doubling time of CIAT899 in 50% TSB media (tryptic soy broth) at 30 ˚C was calculated to be 2.6 hours.

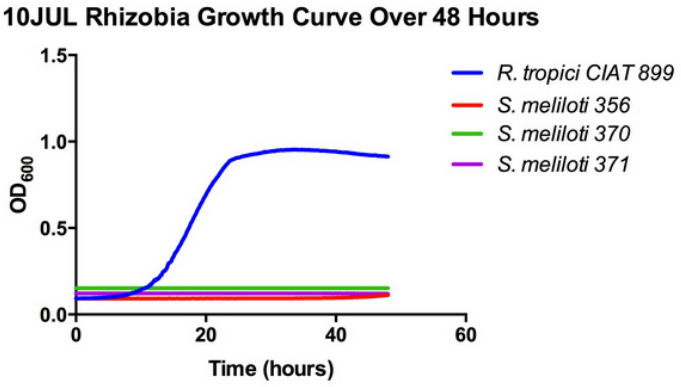
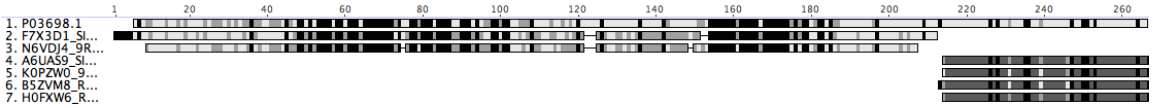


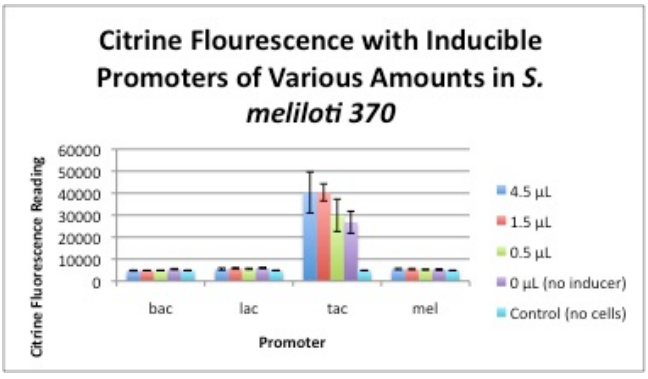
Table 1: Amino acid sequence identities and similarities of homologies to λ-red beta and SPP1 35 found using HMMER.

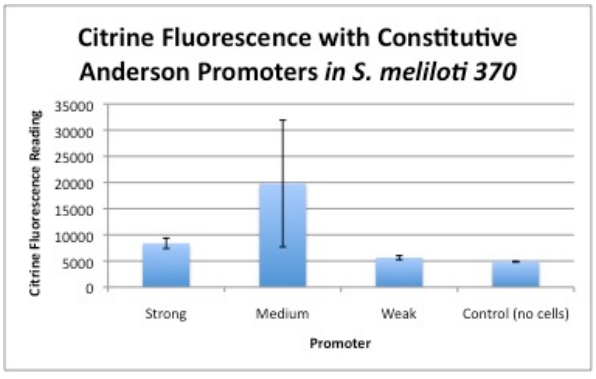
Search sequence	Homologue (UniprotKB Number)	E-value	Query Region	% Base Pair Identity (count)	% Base Pair Similarity (count)
λ-red beta protein (P03698)	F7X3D1	1.50E-36	41-182	51.8 (71)	74.5 (102)
	N6VDJ4	2.30E-42	20-182	51.3 (81)	70.3 (111)
	A6UAS9	2.80E-05	30-85	26.8 (15)	64.3 (36)
			140-222	22.9 (19)	68.7 (57)
	K0PZW0	2.20E-05	28-83	25.0 (14)	66.1 (37)
			131-187	26.3 (15)	71.9 (41)
	B5ZVM8	0.00041	32-82	27.5 (14)	64.7 (33)
			140-189	28.0 (14)	66.0 (33)
	H0FXW6	9.60E-13	36-216	23.8 (43)	68.0 (123)
SPP1 35 protein (Q38143)	L0NDY8	4.90E-12	20-276	30.8 (74)	52.1 (125)
	J0GZ13	1.50E-09	31-239	29.8 (62)	50.5 (105)



Sequence alignment of λ-red beta (P03698) and beta homologues (F7X3D1, N6VDJ4, A6UAS9, K0PZW0, B5ZVM8, H0FXW6).

Figure 5: Characterization of promoter-citrine constructs in *Sinorhizobium meliloti* 1021. The tac promoter (from *E. coli*) appears to be inducible in *S. meliloti*, and the Anderson Medium promoter (obtained as a BioBrick from the Registry of Standard Biological Parts) demonstrates promise as a constitutive promoter in he organism.





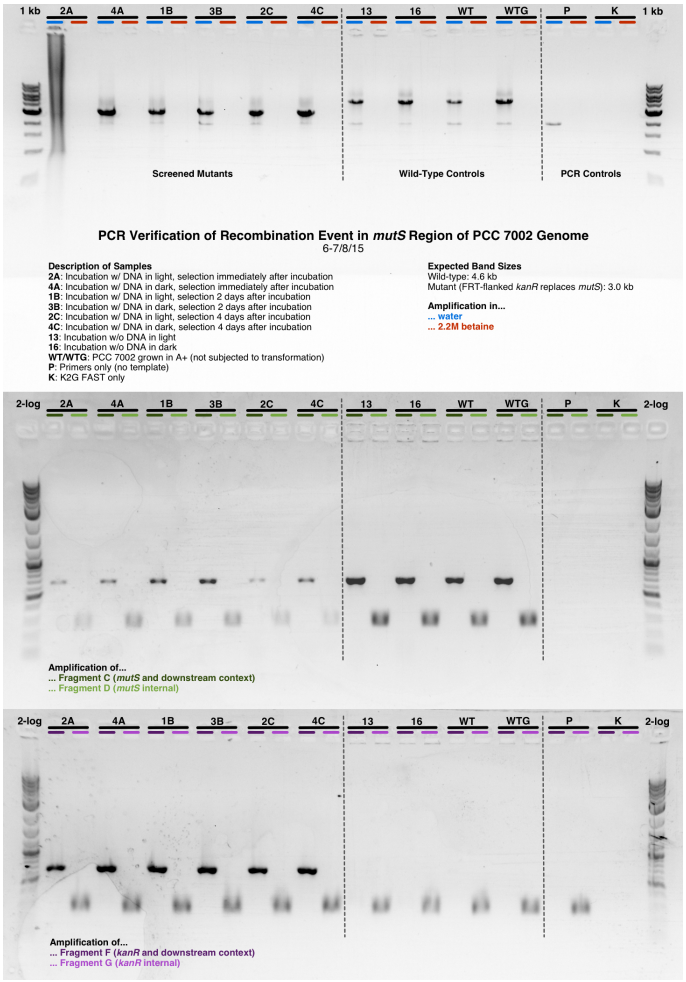


Figure 6: PCR verification of genotypes of PCC7002 samples transformed with linear *kanR* recombination fragments for *mutS* knockout. The middel gel (green) indicates that homozygous Δ*mutS* genotypes are not present in any experimental samples.

Modeling

PCR errors are an important problem for all genetic engineering projects, and particularly relevant towards testing MAGE in a new organism. Reliably tracking the mutations created by a MAGE process should take consideration of the errors generated during PCR, given factors such as sequence length, polymerase type, and number of cycles. We developed an interface to calculate the expected share of mutated sequences given this set of user-defined parameters, which is available on our webpage (http://2015.igem.org/Team:Yale/modeling).

Forecasting Negative Results

Our modeling team developed a PCR error rate estimation program. Originally, our modeling team developed a model based primarily off of binomial distribution. This model had similar results to the error rate calculator hosted by Thermofisher. While not completely correlated, the results were consistently close enough that it is likely these two models had the same underlying basis.

The problem with applying simple binomial distribution to PCR reactions is that the event of copying the DNA is not independent. If an error occurs in an early replication it is propagated through subsequent replications and thus this must be accounted for within the calculation. Original samples of PCR fragments stay in the mixture after they are used as templates in the initial round of replication. Thus, the generation of a PCR fragment matters when replication is occurring. Therefore, while a third cycle fragment's replicated product may have a [1-(0.99)]^3 chance of having a single mutation, there will still be first generation fragments whose replication products will only have a [1-0.99] chance of being mutated. Without taking generations into account, one develops a more pessimistic model than may be accurate. Due to this, we called our original model the Pessimistic Model and the newer model, which we found in literature, the New Model. There are a few other factors that need to be taken into account, such as the chance that a strand of DNA is even replicated at all during a cycle. These concerns are also accounted for in the New Model.

Optimism and Accounting for Sequence History

H. Sharifian 2010 presents a PCR model based on the Galton-Watson process, where an ancestor particle produces a randomly-distributed number of progeny, which repeat this cycle. This branching process can be used to model errors in replication in PCR. This model is based on a previous model by Sun 1995 [1]. We begin with S copies of identical single-stranded DNA fragments. Every cycle, each sequence generates a copy with probability lambda (we assume lambda = 1). With each cycle, PCR errors may occur.

We formulate the error rate per base as mu, and the length of the sequence in

Equation 1

$$
\begin{aligned}
P\{M=m\} &= \sum_{k=0}^{n} P\{M=m|K=k\}P\{K=k\} \\
&= \sum_{k=0}^{n} exp(-k\mu G)\frac{(k\mu G)^m}{m!}\frac{\binom{n}{k}\lambda^k}{(1+\lambda)^n} \\
&= \frac{(\mu G)^m}{m!(1+\lambda)^n}\sum_{k=0}^{n} exp(-k\mu G)\binom{n}{k}k^m\lambda^k \\
&= \frac{(\mu G)^m(1+\lambda e^{-\mu G})^n}{m!(1+\lambda)^n}E\left(Bin\left(n,\frac{\lambda e^{-\mu G}}{\lambda e^{-\mu G}+1}\right)\right)^m
\end{aligned}
$$

question as G. Then, the error rate per sequence is mu*G. We also take into account the fact that later generations of PCR templates have a higher probability of containing errors.

Equation 1 gives the probability of a randomly selected sequence containing m mutations. The parameter k represents the sequence generation, and n represents the number of PCR cycles. The expectation of the binomial is evaluated as the product of the parameters n and (lambda exp(-mu*G)) / (lambda*exp(-mu*G) + 1). We evaluated the above model in a web app which takes in a sequence of DNA, a PCR error rate, and the number of cycles to produce an estimate of the percentage of PCR fragments that contain mutations in the resulting amplified DNA.

References

Cong L, Ann Ran F et al. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* **2013** 339 819-823.

Datsenko KA and Wanner BL. “One-step inactivation of chromosomal genes in *Eschrichia coli* K-12 using PCR products.” *PNAS* **2000** 97(12), 6640-6645.

Frigaard NU, Sakuragi Y, and Bryant DA. “Gene Inactivation in the Cyanobacterium *Synechococcus* sp. PCC 7002 and the Green Sulfur Bacterium *Chlorobium tepidum* Using In Vitro-Made DNA Constructs and Natural Transformation.” In *Photosynthesis Research Protocols*. Methods in Molecular Biology: Humana Press **2004** 274, 325-340.

Hsu P, Lander E & Zhang F. “Development and Applications of CRISPR-Cas9 for Genome Engineering.” *Cell* **2014** 157.

Ludwig M and Bryand DA. “Synechococcus sp. PCC 7002 transcriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions.” *Front. in Microbiol.* **2012** 3(354), 1-14.

Martinez-Romero E, Segovia L et al. "Rhizobium tropici, a Novel Species Nodulating *Phaseolus vulgaris*, L. Beans and *Leucaena* sp. Trees." *Int. J. of System. Biotech.***1991** 41(3), 417-426.

Modrich P and Lahue R. “Mismatch repair in replication fidelity, genetic recombination, and cancer biology.” *Annual Reviews* **1996** 65, 101-133.

Ruffing A. "Improved Free Fatty Acid Production in Cyanobacteria with Synechococcus sp. PCC 7002 as Host." *Front. Bioeng. Biotechnol.* **2014** 2.

Schweizer HP. “Applications of the Saccharomyces cerevisiae Flp-FRT System in Bacterial Genetics” *J. Mol. Bicrobiol. Biotechnol*. **2003** 5, 67-77.

Sharifian, Hoda. Errors induced during PCR amplification. Swiss Federal Institute of Technology, Department of Computer Science (2010). http://dx.doi.org/10.3929/ethz-a-006088024

Tan X, Liang F, Cai K, and Lu X. “Application of the FLP/FRT recombination system in cyanobacteria for construction of markerless mutants.” *Appl. Microbiol. Biotechnol*. **2013** 97, 6737-6382.

Acknowledgements

All iGEM team members, as well as advisors Natalie Ma, Dr. Farren Isaacs, and Dr. Stephen Dellaporta, participated in the formation of the idea for this project and the research required to design and carry out this work. Ben Akhuetie-Oni, Edward Barbieri, Ryan Gallagher, Adrian Haimovich, Paul Muir, Jaymin Patel, Corey Perez (Yale Crawford Lab), and Ariel Hernandez-Leyva helped with running lab experiments as well as lending resources. Natalie Ma, in addition to aiding with experiments, was also instrumental in keeping the team on track by helping us set daily and weekly goals. The Jordan Peccia Lab at Yale University provided us with the *Synechococcus elongatus* UTEX 2973 cyanobacteria strain and gave us advice on growing and culturing cyanobacteria. The Handelsman Lab and the Jacobs-Wagner Lab at Yale University provided us with the *Rhizobium tropici* strain CIAT 899 and *Sinorhizobium meliloti* 1021 strains 356, 370, and 371, respectively.

The corporate sponsors for this project were IDT, AmericanBio, New England Biolabs, KAPA Biosystems, Geneious, and MathWorks. We would also like to thank the following Yale entities for their financial support: Yale West Campus; Yale Department of Molecular, Cellular, and Developmental Biology; Yale Department of Biomedical Engineering; Yale Department of Molecular Biophysics and Biochemistry; Yale University Provost’s Office; Yale Uniergraduate Organizations Committee; and the Yale Freshman Summer Fellowship and STARS Programs.